Standard errors of heritability estimates

Iysik.com

Standard errors of heritability estimates

Author: Stefan McKinnon Edwards
Date: September 2017

PDF version with equations correctly formatted. dmu_heritabilities.pdf

Cite as:

> Stefan McKinnon Edwards (2017) *Standard errors of heritability estimates*
> Online (http://www.iysik.com/dmu/heritabilities [http://www.iysik.com/dmu/heritabilities])

```
@online{h2se,
  author = {Stefan McKinnon Edwards},
  urldate = {2017-09-21},
  url = {http://www.iysik.com/dmu/heritabilities},
  title = {Standard errors of heritability estimates},
  year = {2017}
}
```

The heritability is calculated as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$. DMU (Madsen et al., 1994, Madsen & Jensen, 2000) presents the uncertainties of these estimates as the asymptotic standard errors se_g and se_e, with asymptotic correlation matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. The standard error of the heritability estimate is then:

$$s.e.(h^2) = \sqrt{\left[\left(\frac{\partial h^2}{\partial \sigma_g^2}\right)^2 (se_g)^2 + \left(\frac{\partial h^2}{\partial \sigma_e^2}\right)^2 (se_e)^2 + 2\left(\frac{\partial h^2}{\partial \sigma_g^2}\right)\left(\frac{\partial h^2}{\partial \sigma_e^2}\right)\rho_{g,e} se_g se_e\right]}$$

The partial derivatives are given as

- $\frac{\partial h^2}{\partial \sigma_g^2} = \frac{\sigma_e^2}{(\sigma_g^2+\sigma_e^2)^2}$
- $\frac{\partial h^2}{\partial \sigma_e^2} = \frac{-\sigma_g^2}{(\sigma_g^2+\sigma_e^2)^2}$
- $\rho_{g,e} se_g se_e = cov\left(\hat{\sigma}_g^2, \hat{\sigma}_e^2\right)$

Derivation

The details of these calculations were presented as notes in personal communications from DMU's authors, Per Madsen. I am retracing them here to validate the above expressions.

The heritability, and similar genetic parameters, are non-linear transformations of the vector of variance components $\Theta = (\sigma_1^2, \sigma^2, \ldots, \sigma_t^2)$. I.e. we can write it more generally as $h^2 = f(\Theta) = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2}$, for a subset of elements of Θ. If $f(\Theta)$ is differentiable and continuous around the estimated $\hat{\Theta}$, then we can perform Taylor series expansion of the transformation. This provides a tangible expression of the pairwise, conditional co-variances of the transformation. As we are interested in the variance of the transformation, we will take steps to obtain the expectation of the squared deviation from the mean. A key here is that $E[\Theta|y] = \hat{\Theta}$, when $\hat{\Theta}$ is the maximizer of the likelihood function. The final expression is found in equation (6). But first, the Taylor series expansion:

$$f(\Theta) = f(\hat{\Theta}) + \sum_{i=1}^{t} \frac{\partial f}{\partial \sigma_i^2}\Big|_{\sigma_i^2=\hat{\sigma}_i^2} (\sigma_i^2 - \hat{\sigma}_i^2) \qquad (1)$$

Taking expectations on both sides we get:

$$E[f(\Theta)] = f(\hat{\Theta}) \qquad (2)$$

Subtracting (2) from (1), the right hand side of (2) cancels the first term in (1), and the left hand side is the deviation from the expectation, as given in the following expression:

$$f(\Theta) - E[f(\Theta)] = \sum_{i=1}^{t} \frac{\partial f}{\partial \sigma_i^2}\Big|_{\sigma_i^2=\hat{\sigma}_i^2} (\sigma_i^2 - \hat{\sigma}_i^2) \qquad (3)$$

Final step is squaring and taking the expectation. The left-hand side of (3) becomes

$$E(f(\Theta) - E[f(\Theta)])^2 = Var(f(\Theta)) \qquad (4)$$

i.e. the variation of the non-linear transformation. The left-hand side of (3) becomes a tangible expression:

$$\left[\sum_{i=1}^{t} \frac{\partial f}{\partial \sigma_i^2}\Big|_{\sigma_i^2=\hat{\sigma}_i^2} (\sigma_i^2 - \hat{\sigma}_i^2)\right]^2 = \sum_{i=1}^{t} \left(\frac{\partial f}{\partial \sigma_i^2}\Big|_{\sigma_i^2=\hat{\sigma}_i^2}\right)^2 (\sigma_i^2 - \hat{\sigma}_i^2)^2 \qquad (5)$$

Remembering that $E(\sigma_i^2) = \hat{\sigma}_i^2$ (when $\hat{\sigma}_i^2$ maximizes the likelihood), we can recognize $Var(\sigma_i^2) = (\sigma_i^2 - \hat{\sigma}_i^2)^2 = \sigma^2_{\sigma_i^2}$, i.e. the variance of the variance component.

Similarly, $Cov(\sigma_i^2, \sigma_j^2) = (\sigma_i^2 - \hat{\sigma}_i^2)(\sigma_j^2 - \hat{\sigma}_j^2) = \sigma_{\sigma_i^2,\sigma_j^2}$. Thus the Taylor series expanded estimate of the variance of a non-linear transformation of the variance component becomes

$$Var\left(f(\Theta)\right) = \sum_{i=1}^{t} \left(\frac{\partial f}{\partial \sigma_i^2}\Big|_{\sigma_i^2=\hat{\sigma}_i^2}\right)^2 \sigma^2_{\sigma_i^2} + 2\sum_{i=1}^{t}\sum_{j=i+1}^{t} \frac{\partial f}{\partial \sigma_i^2} \cdot \frac{\partial f}{\partial \sigma_j^2} \cdot \sigma_{\sigma_i^2,\sigma_j^2} \qquad (6)$$

When using the AI-REML algorithm (Gilour et al., 1995; Johnson & Thompson, 1995; Jensen et al., 1997) in DMU for estimating variance components, the inverse of the average information matrix, $I_O(\Theta)$, contains the asymptotic variance-covariance of Θ. I.e. $Var(\Theta) = I_O^{-1}(\Theta)$. Asymptotic variances of variance components, $\sigma^2_{\sigma^2_i}$, are on the diagonal, and co-variances, $\sigma_{\sigma^2_i,\sigma^2_j}$, are found on the off-diagonal. $\sqrt{\sigma^2_{\sigma^2_i}}$ provides the asymptotic standard error.

With DMU

I provide here some expressions for common genetic parameters. The standard error of the estimate is reported as $se_{h^2} = \sqrt{Var(h^2)}$.

As DMU in the output log returns asymptotic *standard errors* of the estimated variance components, and asymptotic *correlations* between variance components, some re-formulation is required. The partial derivatives (e.g. $\partial h^2 / \partial \sigma^2_g$) can be calculated with values plugged in directly from DMU, whereas the final estimate cannot. Recall that the correlation is no more than the covariance scaled by the respective squared variances:

$$Cor(\sigma^2_i, \sigma^2_j) = \frac{Cov(\sigma^2_i, \sigma^2_j)}{\sqrt{Var(\sigma^2_i)}\sqrt{Var(\sigma^2_j)}} = \frac{\sigma_{\sigma^2_i,\sigma^2_j}}{se_{\sigma^2_i} \cdot se_{\sigma^2_j}}$$

with the product in the denominator ($se_{\sigma^2_i}$ and $se_{\sigma^2_j}$) simply being the asymptotic standard errors. Hence, using the output from DMU, the values needed can be calculated as:

$$\sigma_{\sigma^2_i,\sigma^2_j} = Cor(\sigma^2_i, \sigma^2_j) \cdot se_{\sigma^2_i} \cdot se_{\sigma^2_j} \qquad \text{and} \qquad \sigma^2_{\sigma^2_i} = \left(se_{\sigma^2_i}\right)^2$$

Some genetic parameters

Heritability (narrow-sense)

For

$$h^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e},$$

then

$$\frac{\partial h^2}{\partial \sigma^2_g} = \frac{\sigma^2_e}{(\sigma^2_g + \sigma^2_e)^2}$$

and

$$\frac{\partial h^2}{\partial \sigma^2_e} = \frac{-\sigma^2_g}{(\sigma^2_g + \sigma^2_e)^2}$$

Note: There really *is* a negative sign in the latter. Note: Use the quotient rule for deriving partial derivations.

$$Var(h^2) = \left(\frac{\partial h^2}{\partial \sigma^2_g}\right)^2 \sigma^2_{\sigma^2_g} + \left(\frac{\partial h^2}{\partial \sigma^2_e}\right)^2 \sigma^2_{\sigma^2_e} + 2\left(\frac{\partial h^2}{\partial \sigma^2_g}\right)\left(\frac{\partial h^2}{\partial \sigma^2_e}\right)\sigma_{\sigma^2_g,\sigma^2_e}$$

Hertiability (total)

For

$$h^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_m + \sigma^2_e},$$

then

$$\frac{\partial h^2}{\partial \sigma^2_g} = \frac{\sigma^2_m + \sigma^2_e}{(\sigma^2_g + \sigma^2_m + \sigma^2_e)^2}$$

and

$$\frac{\partial h^2}{\partial \sigma^2_m} = \frac{\partial h^2}{\partial \sigma^2_e} = \frac{-\sigma^2_g}{(\sigma^2_g + \sigma^2_m + \sigma^2_e)^2}.$$

Therefore,

$$\begin{aligned} Var(h^2) = &\left(\frac{\partial h^2}{\partial \sigma^2_g}\right)^2 \sigma^2_{\sigma^2_g} + \left(\frac{\partial h^2}{\partial \sigma^2_m}\right)^2 \sigma^2_{\sigma^2_m} + \left(\frac{\partial h^2}{\partial \sigma^2_e}\right)^2 \sigma^2_{\sigma^2_e} + \\ &2\left(\frac{\partial h^2}{\partial \sigma^2_g}\right)\left(\frac{\partial h^2}{\partial \sigma^2_e}\right)\sigma_{\sigma^2_g,\sigma^2_e} + 2\left(\frac{\partial h^2}{\partial \sigma^2_g}\right)\left(\frac{\partial h^2}{\partial \sigma^2_m}\right)\sigma_{\sigma^2_g,\sigma^2_m} + \\ &2\left(\frac{\partial h^2}{\partial \sigma^2_m}\right)\left(\frac{\partial h^2}{\partial \sigma^2_e}\right)\sigma_{\sigma^2_m,\sigma^2_e} \end{aligned}$$

Genetic correlation

The genetic correlation can be estimated using a bivariate model, such as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

where $Var(g_1) = G\sigma_1^2$ and $Var(g_2) = G\sigma_2^2$ with variance-covariance matrix G for both random effects, and $Cov(g_1, g_2) = \sigma_{1,2}$ as the genetic covariance between two traits y_1 and y_2. In this context, $\sigma_{1,2}$ is no different from σ_1^2, as they are both variance components that are estimated by e.g. AI-REML.

For genetic *correlation*

$$r = \frac{\sigma_{1,2}}{\sqrt{\sigma_1^2 \cdot \sigma_2^2}}$$

then

$$\frac{\partial r}{\partial \sigma_{1,2}} = \frac{1}{\sqrt{\sigma_1^2 \cdot \sigma_2^2}},$$

$$\frac{\partial r}{\partial \sigma_1^2} = \frac{-\sigma_{1,2}}{2 \cdot \left(\sqrt{\sigma_1^2}\right) \cdot \left(\sigma_2^2\right)^{\frac{3}{2}}},$$

and

$$\frac{\partial r}{\partial \sigma_2^2} = \frac{-\sigma_{1,2}}{2 \cdot \left(\sqrt{\sigma_2^2}\right) \cdot \left(\sigma_1^2\right)^{\frac{3}{2}}}.$$

Therefore,

$$\begin{aligned} Var(r) = &\left(\frac{\partial r}{\partial \sigma_1^2}\right)^2 \sigma^2_{\sigma_1^2} + \left(\frac{\partial r}{\partial \sigma_2^2}\right)^2 \sigma^2_{\sigma_2^2} + \left(\frac{\partial r}{\partial \sigma_{1,2}}\right)^2 \sigma^2_{\sigma_{1,2}} + \\ &2\left(\frac{\partial r}{\partial \sigma_1^2}\right)\left(\frac{\partial r}{\partial \sigma_{1,2}}\right) \sigma_{\sigma_1^2,\sigma_{1,2}} + 2\left(\frac{\partial r}{\partial \sigma_1^2}\right)\left(\frac{\partial r}{\partial \sigma_2^2}\right) \sigma_{\sigma_1^2,\sigma_2^2} + \\ &2\left(\frac{\partial r}{\partial \sigma_2^2}\right)\left(\frac{\partial r}{\partial \sigma_{1,2}}\right) \sigma_{\sigma_2^2,\sigma_{1,2}} \end{aligned}$$

References

To cite the software DMU, refer to *Madsen et al., 1994* and *Madsen & Jensen, 2000*. When using the AI-REML algorithm for estimating variance components (with `dmuai`) also refer to *Gilmour et al., 1995* and *Johnson & Thompson, 1995*, and *Jensen et al., 1997*.

1. Gilmour, A. R., R. Thompson, and B. R. Cullis, 1995 Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. Biometrics 51: 1440–1450.
2. Jensen, J., E. A. Mantysaari, P. Madsen, and R. Thompson, 1997 Residual Maximum Likelihood Estimation of (Co) Variance Components in Multivariate Mixed Linear Models using Average Information. J. Indian Soc. Agric. Stat. 49: 215–236.
3. Johnson, D. L., and R. Thompson, 1995 Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information. J. Dairy Sci. 78: 449–456.
4. Madsen P, Jensen J, Thompson R. Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information. 5th WCGALP. Guelph, Canada; 1994. p. 455–62.
5. Madsen P, Jensen J. A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, release 5.1 [Internet]. Tjele, Denmark; 2000. p. 32. Available from: http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf [http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf]

dmu.bib

```
@inproceedings{Madsen1994,
address = {Guelph, Canada},
author = {Madsen, Per and Jensen, Just and Thompson, Robin},
booktitle = {5th WCGALP},
number = {1977},
pages = {455--462},
title = {{Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information}},
year = {1994}
}
@article{Gilmour1995,
author = {Gilmour, Arthur R. and Thompson, Robin and Cullis, Brian R.},
issn = {0006341X},
journal = {Biometrics},
number = {4},
pages = {1440--1450},
publisher = {International Biometric Society},
title = {{Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models}},
url = {http://www.jstor.org/stable/2533274},
volume = {51},
year = {1995}
}
@article{Jensen1997,
author = {Jensen, Just and Mantysaari, Esa A. and Madsen, Per and Thompson, Robin},
journal = {Journal of Indian Society of Agricultural Statistics},
keywords = {AI-REML,In binder: LMM,REML},
mendeley-groups = {Papers/Thesis,Software,Papers/2017 Canine Hip Dysplasia},
mendeley-tags = {AI-REML,In binder: LMM,REML},
pages = {215--236},
title = {{Residual Maximum Likelihood Estimation of (Co) Variance Components in Multivariate Mixed Linear Models using Average Information}},
url = {http://isas.org.in/jisas/jsp/abstract.jsp?title=Residual Maximum likelihood Estimation of (Co) Variance Components in Multivariate Mixed Linear Models Using A
volume = {49},
year = {1997}
}
@article{Johnson1995,
author = {Johnson, D.L. and Thompson, Robin},
doi = {10.3168/jds.S0022-0302(95)76654-1},
file = {:C$\backslash$:/Users/shojedw/AppData/Local/Mendeley Ltd./Mendeley Desktop/Downloaded/Johnson, Thompson - 1995 - Restricted Maximum Likelihood Estimation of
issn = {00220302},
journal = {Journal of Dairy Science},
keywords = {AI-REML,Per's references},
mendeley-groups = {Papers/Thesis,Zotero - Zotero Library,Genomic selection and prediction,Software,Papers/2017 CVAT,Papers/2017 Canine Hip Dysplasia,Papers/2015 Part
```

```
mendeley-tags = {Per's references},
month = {feb},
number = {2},
pages = {449--456},
publisher = {Elsevier},
title = {{Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information}},
url = {http://www.journalofdairyscience.org/article/S0022-0302(95)76654-1/abstract},
volume = {78},
year = {1995}
}
@book{DMU5.1,
address = {Tjele, Denmark},
author = {Madsen, Per and Jensen, Just},
keywords = {AI-REML,DMU},
mendeley-tags = {DMU},
pages = {32},
title = {{A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, release 5.1}},
url = {http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6{\_}guide.5.2.pdf},
year = {2000}
}
```
